Understanding the Time-Course of an Intervention's Mechanisms: A Framework for Improving Experiments and Evaluations

Shannon J. Linning*	**Kate Bowers**	**John E. Eck**
Assistant Professor & Corresponding Author* Department of Criminal Justice and Criminology Washington State University PO Box 644872 Pullman, WA, USA 99164-4872 shannon.linning@wsu.edu (509) 335-8428	Professor Jill Dando Institute of Security and Crime Science University College London 35 Tavistock Square London, UK WC1H 9EZ kate.bowers@ucl.ac.uk	Professor School of Criminal Justice University of Cincinnati P.O. Box 210389 Cincinnati, OH 45211-0389 john.eck@uc.edu

ABSTRACT

Objectives: The crime prevention evaluation literature has identified several potential side-effects of interventions. These often-unintended consequences occur at different stages of prevention processes, including before official start dates. They can improve or reduce intervention impacts. Evaluations using before-and-after designs with or without controls can fail to identify these effects. We describe a longitudinal framework to guide the design and evaluation of interventions that can account for these side-effects when causal mechanisms are better understood.

Methods: Our time-course framework provides a comprehensive assessment of the prevention process. Using place-based examples as illustrations, it builds on previously identified temporal benefits and backfires—such as anticipatory benefits, residual deterrence, and initial backfire—that have never been systematically organized into a single framework. We show how our framework can be incorporated into the EMMIE framework for assessing prevention utility.

Results: The proposed time-course framework links together all temporal effects, their underlying mechanisms, and shows how they can vary by context.

Conclusions: The framework suggest that considering all decisions within these timelines will be more cost-effective and produce greater crime reductions in the long-run. By considering the mechanisms that can be triggered at various points in an intervention's time-course, we can better design experiments to test them and generate stronger evaluations of programs.

Keywords: Crime prevention policy; EMMIE framework; Initial backfire; Intervention time-course; Program evaluation

INTRODUCTION

Policy researchers argue that the most effective policies are evidence-driven. However, determining the effectiveness of a strategy is not easy. Considerable attention has been paid to how one synthesizes evidence from several studies and to adding more policy-relevant information (see discussions of EMMIE in Johnson et al. 2015). Here, we focus on a different, though related, set of problems: that prevention interventions can create effects that occur at different times—before, during, and after—because the intervention triggers different mechanisms.

Several of these can be found in the crime prevention literature. In addition to the direct and anticipated effects, several unplanned effects have been discovered by researchers. Both crime displacement and diffusion of crime control benefits are the most well-known unanticipated spatial effects (Clarke & Weisburd 1994; Guerette & Bowers 2009). In addition to these, several unanticipated temporal effects have since been introduced. These include anticipatory benefits (Smith et al. 2002), residual deterrence (Sherman 1990), and initial backfire (Linning and Eck 2018). This paper focuses on these temporal effects that may occur before, during, and after implementation.

Because these temporal effects were discovered individually, they are usually considered in isolation. Here, we link these effects into a single framework. Nearly all interventions trigger multiple mechanisms that can produce intended or unplanned effects. However, the timing of them is not often considered. There is much we do not know. Can mechanisms only be triggered at certain times (i.e., before, during, or after an intervention)? Are certain mechanisms dependent on prior ones or can they work freely? And are the effects produced by these mechanisms the same in different contexts? Although tying temporal effects together might seem obvious, it has never been suggested in the policy literature. A unified approach to these temporal effects can assist in the development of more effective interventions. Interventions may be over- or under-estimated if practitioners and evaluators do not consider an intervention's full time-course. Thus, this is a methodological paper with the goal of improving experiments and evaluations. Use of our framework will lead to better estimates of an intervention's effectiveness and will reduce bias in systematic reviews.

We have organized our paper as follows. The first section summarizes the research on these temporal effects. In the second section, we discuss the temporal mechanisms and how they work in various contexts. Finally, we provide suggestions on how our framework can improve experiments and evaluations.

CONSOLIDATING THE EFFECTS ON A TIMELINE

Varying outcomes can generally be expected across three time points: before, during, or after an intervention. It is possible to experience benefits or backfires at any of these stages. However, researchers did not discover various temporal effects in the order in which they typically occur during the time-course of an intervention. They emerged from independent study of often seemingly anomalous consequences of interventions. We discuss these effects in the order which they would occur during an intervention, rather in the order in which they appear in the literature.

Benefits before an intervention. Smith et al. (2002) found many examples where crime actually declined prior to the implementation of certain interventions, which they named *anticipatory benefits*. Anticipatory benefits occur for several reasons. For instance, offenders may become aware of upcoming prevention strategies via media campaigns. This leads them to reconsider their likelihood of getting caught and often suppresses their offending even before the intervention takes effect. Another example is that they believe the intervention has begun before it actually has. Then once the strategy begins, it continues this suppression of offending.

Benefits during and after an intervention. After discussions of initial deterrence surfaced (Ross 1981), Sherman (1990) introduced the concept of residual deterrence. He observed sustained crime reductions following the end of successful police crackdowns. Crackdowns —a sudden temporary influx of police in a small area—create a period of intended initial deterrence whereby offenders refrain from crime because their risk of detection increases. Sherman (1990) observed that this reduction in crime persisted after a crackdown ended. He believed this occurred because offenders had limited knowledge of police actions, so they were careful about their behavior. He noted that offenders' concern about police presence "[decayed] slowly" (Sherman 1990: 3). Thus, police departments could actually reduce their resources at these places temporarily while continuing to reap the "bonus" effect of the crackdown.

Backfires at all intervention stages. Not all unanticipated temporal effects are beneficial. Linning and Eck (2018) proposed the concepts of weak intervention, initial, and anticipatory backfire. The first argues that if an intervention incorrectly or insufficiently disrupts the opportunities for crime, it can actually create new opportunities for offending. They then extended their ideas by proposing temporal extensions called: initial and anticipatory backfire. This suggests that while some interventions can backfire completely, others may only backfire temporarily.

LINKING THE TEMPORAL EFFECTS

A single time-course intervention framework unites these effects. We focus on six general outcomes that could emerge over the course of an intervention: 1) anticipatory benefits, 2) anticipatory backfire, 3) initial deterrence, 4) initial backfire, 5) residual deterrence, and 6) residual backfire. Each of these effects are driven by offenders' perception of risks and opportunities. This implies that more than one of these effects could transpire from the same intervention. To be clear, effects will be observed at three stages—before (anticipatory), during (initial), and after (residual)—but which effects occur will depend on which mechanisms the intervention triggers.

<INSERT FIGURE 1 ABOUT HERE>

Before we consider the time-course of an intervention, we must first establish an offending timeline. Figure 1 shows a hypothetical offending timeline. Once offenders have identified suitable crime opportunities, they begin offending. It is not until opportunities are disrupted that offending will change. At the bottom of figure 1, we show the implementation timeline. This begins when a crime problem is detected. We assume this is some time after the offenders have started operating. From here on we consider how events on the implementation timeline influence the offenders' timeline. The first potential stimulus that might cause offenders to change will be any announcement of an intervention. This could be intentional (e.g., a press release) or it could be fortuitous (e.g., technicians installing lights). Offenders may adjust their cost-benefit thinking, ideally leading to a decline in crime because of increased uncertainty. But it is possible that anticipation of prevention might stimulate increased offending. For example, prior to a political protest, the mayor declares he will have zero tolerance for misbehavior. Angered by the perceived insults, potential protesters initiate misbehavior even before the planned protest; an *anticipatory backfire*. Moving from planning to implementation, offenders may err on the side of caution and refrain from crime when uncertainty of risk is high; *initial deterrence*. Here too backfire is theoretically possible, particularly when the intervention is weak. Finally, at the end of the intervention we may observe *residual deterrence or backfire*.

How sustainable are the effects after the so-called "residual" period? Interventions that have sustained positive effects over time are desirable. Sustainability probably depends in part on the type of environment within which the intervention is implemented. We can use the EMMIE framework to understand sustainability. The EMMIE framework (Johnson et al. 2015) suggests there are multiple dimensions that should be considered to understand a particular outcome. These dimensions include: *effect* size and direction, *mechanism*/mediators (how an intervention works), *moderators*/contexts (the conditions necessary for the intervention to work), *implementation* success or failure (required resources and structures), and *economic* costs (consideration of financial value).

To illustrate, consider an example of burglaries in a residential neighborhood. We will assume that an increase in residential burglaries will occur both temporally and geographically close to the first home that is victimized: namely, near repeats (i.e., Bowers and Johnson 2005; Johnson and Bowers 2004). This means that once a suitable first target is selected by an offender, other homes within its vicinity also become the targets for subsequent burglaries. Let us also assume that these dwellings are attached (i.e., row homes) with alley ways behind them. We will consider two contrasting potential interventions—a police crackdown and alley-gating—designed to reduce these burglaries.

<INSERT FIGURE 2 ABOUT HERE>

Figure 2 provides an intervention time-course for a police crackdown strategy that also highlights all EMMIE dimensions. For the sake of argument, we assume that police make no pre-intervention announcement of their intent. Thus, changes in offending are only expected once the intervention actually begins. This means that no anticipatory benefits or backfires occur. Offenders only change their behavior once the police are seen patrolling the area (i.e., visibility).

And they will only alter their behavior if they see the officers as a legitimate source of authority (i.e., credibility). Declines in crime via initial and residual deterrence would occur during the intervention period and persist for some time after patrols have stopped. But the sustainability of this burglary decline is unclear. Unless the police return to this area after the residual deterrence period, increases in crime would be expected. That is, as soon as offenders learn “that it is once again ‘safe’ to offend”, they will resume such activities (Sherman 1990: 10). Thus, to temporally sustain desired effects, the police would need to continue putting resources into the intervention and their patrols (seen on the far right of the timeline). This would require a ‘pulse’ policing style (Sherman 1990).

In this intervention, the three main mechanisms function sequentially. Specifically, the uncertainty mechanism is dependent on the visibility and credibility mechanisms. A sustained decline in offending post-intervention is caused by offenders’ uncertainty of whether the police will resume their crackdown. However, this can only occur if the police are perceived as legitimate sources of authority and had a visible presence in the area to begin with. Thus, the uncertainty mechanism can only be activated if the visibility and credibility mechanisms were triggered first. This is an example of a small *mechanism cascade* whereby achieving desired results from one mechanism is dependent on the successful triggering of an earlier mechanism. Gambetta (1998) provides a similar explanation for mechanisms interacting with one another that form “concatenations of mechanisms” (p. 105).

Figure 3 provides a time-course but using a different prevention strategy. Here, practitioners implement alley-gating. Alley-gates are devices commonly used in the United Kingdom to restrict access to alleyways. Workers install a barricade at both ends of an alley, with a lockable gate in each barricade. Once installed, only residents and city workers are given keys, making passage through these areas difficult to non-residents (see Bowers et al. 2004a).

<INSERT FIGURE 3 ABOUT HERE>

In the case of alley-gating, anticipatory benefits would be expected because of "preparation-disruption effects" (Smith et al. 2002: 78). The installation of the alley-gates is clearly visible to anyone in the neighborhood and they are seldom installed without consulting property owners. Even before the gates are present, offenders would observe crews working in these areas as their sites are measured, prepared, and fitted. The workers' presence would increase offenders' risk of detection thus deterring them from targeting these areas before the intervention technically begins. Once the gates are installed, one would expect the reduction in crime to persist. That is, the desired effects will continue even though the reasons for this crime decline have changed. Here, an increased effort mechanism is triggered (see Sidebottom et al. 2017). It should also be noted that this mechanism is not dependent on whether the anticipatory mechanism is triggered. In other words, the mechanism cascade effect does not apply because the increased effort mechanism operates independently once gate installation is complete.

Moreover, no residual deterrence would take place. As per Sherman's (1990) explanation, residual deterrence involves the decay of perceived risk leading to an eventual return to offending. But given the increased effort mechanism and that the gates are permanent structures, there would be no such decay. Moreover, the permanent nature of the gates will reap desired effects for many years thereafter – assuming the gates are properly maintained. Another advantage of using alley-gates is that the main economic inputs are only required for a short and limited amount of time. This does not mean their effects will last forever: over time the gates will deteriorate, and residents may become less attentive to locking them.

Next we turn to the questions that this framework highlights: under what circumstances do these effects occur? Why do some temporal effects materialize for a given intervention and not others? Are they always expected? Can we usefully manipulate them?

THE MECHANISMS OF TEMPORAL EFFECTS

If we can understand what the conditions are that create a particular temporal effect, can we manipulate the likelihood of them occurring more intentionally? This question brings up a host of practical issues. Both practitioner accounts and the general crime reduction research literature tend to describe interventions in a simple sense: we did X using Y resources, in attempting to reduce crime problem Z. But what other causal processes did the X set off or inhibit leading to outcomes beyond the intended reduction in Z? The process of implementing intervention X may have an assortment of impacts on conditions the intervention was not explicitly designed to address.

The language of realistic evaluation provides some assistance (Pawson & Tilley 1997). Causal processes here are termed ‘mechanisms’ and the idea is to identify the necessary conditions under which a certain mechanism will be activated to produce a particular outcome. Sadly, it appears that the main mechanisms of interventions are still both poorly articulated and poorly evidenced, as it apparent through a recent examination of many systematic reviews in crime reduction (Thompson et al. 2019).

To illustrate the utility of examining mechanisms, we will continue to examine residential burglary. We might ask how burglar alarms protect homes against burglary. To do this, we need to track the logical steps between the output (a fitted burglar alarm) and the desired effect or outcome (burglary reduction). An obvious chain of events is that a would-be burglar sees the alarm box, perceives an increased risk of apprehension, and decides that it is too risky to proceed further. In crime reduction, there are often alternative mechanisms that can be mapped that depend on specific conditions. Hence, if the alarm is not observable from the outside, a more likely chain of events would be that the offence is attempted and the alarm is triggered, thus scaring off the offender. This rather simple example demonstrates the need to consider the interactions between mechanisms and conditions. Pawson and Tilley (1997) refer to these as Context-Mechanism-Outcome Configurations (CMOC).

In a similar way, it is obvious that 'unintended outcomes', 'spin-offs' and/or 'temporal effects' will have their own series of CMOCs. Prior research has begun to address this. Smith et al (2002) describe ten different mechanisms by which an anticipatory benefit is possible. It is apparent that the degree to which these alternatives are likely to occur will depend on the conditions under which implementation occurs. We can take Smith et al.'s (2012; 78) "preparation-disruption effects" mechanism as an example. If an intervention (such as alley-gating) requires consultation with homeowners, then the implementing agency will need to contact all residents. They might do this in several ways. Face-to-face consultation involving knocking on resident's doors is likely to increase natural surveillance, whereas a posted letter consultation will not do this. So, preparation-disruption is likely in one condition, but not the other. Likewise, both weak intervention backfire (Linning and Eck 2018) and residual deterrence (Sherman 1990) are in fact mechanistic explanations for expecting certain outcomes at particular points in time that, crucially, depend on particular conditions being activated. Weak intervention backfire occurs when the dose provided is insufficient to allow the main mechanism to kick-in. Likewise, residual deterrence may be more likely to occur when police crackdowns are short but unpredictable in duration because offenders have difficulty adjusting to the presence or absence of police. Again, the temporal effect had specific conditions attached.

So, certain actions and conditions can make temporal effects more likely. To encourage the positive ones, where doing so would lead to extra benefit, we need to understand that they are caused by ‘sub-mechanisms’. By this we mean extra chains of events that occur alongside, and interact with, the more explicitly intended crime reduction mechanism of the intervention. Mapping these is obviously an enormous challenge, but we can imagine a crime prevention intervention to create a cascade of events, some of which are direct consequences of the intervention and some of which are side effects. Another interesting implication of considering mechanism cascades is the time period over which these interactions might occur. The entire process—from the first effect to the last—will likely be longer for certain interventions than others. It seems probable that personnel-based interventions will have shorter cascade periods than physical-design ones.

Pawson and Tilley (1997) suggest that the explanation of what works to reduce social problems should contain enough of a detailed explanation of how it works, and the contingencies required to make it work, to be useful and useable by practitioners. In essence, they suggest that reports of evaluation findings should centre on the external validity of the program; the degree to which it is likely that the activity could be transplanted in another situation and still have a fighting chance of gaining the same outcome. To guard against unintended effects, or to fully articulate what might be expected elsewhere, such explanations ought to consistently document sub-mechanisms (and their initiating conditions) as well as main ones. As Eck (2010) suggests, it is important that studies either describe interventions in sufficient detail so others could faithfully implement them, or clearly state that there is a lack of relevant information, so that policy makers can most effectively share this information or acknowledge the need for further documentation.

There may be strategies that promote the more desirable temporal effects across more than one situation. Bowers and Johnson (2005) suggest that publicity might be a cheap and effective method of encouraging both anticipatory benefit and residual deterrence. Evidence from a large-scale evaluation of burglary schemes in England suggested that intervention-specific publicity (i.e., carefully targeted publicity) coincided with anticipatory drops in crime levels. Whilst this evidence is not sufficiently exhaustive to rule out alternative explanations, it does make sense mechanistically. Through the publicity mechanism (Smith et al. 2002), announcing an intervention that is either imminent or has recently been implemented can give potential offenders cause to believe that they are at a higher risk of apprehension if they attempted an offence. Publicity tends to be very cheap compared to other measures; using it to draw out further reduction whilst only modestly increasing cost should improve overall cost-effectiveness. It is important to note however, that intervention-based publicity is far less likely to be as effective in the absence of any on the ground activity and so should be administered as a sub-mechanism. Equally, a concerted attempt to inject dosage upfront in an intervention's lifetime, might not even involve increasing resources at all, merely a concerted re-distribution of them. If this is done sufficiently, it could lead to the blocking of the sub-mechanism that leads to initial backfire.

IMPROVING EXPERIMENTS AND EVALUATIONS

A better understanding of *when* we expect to trigger mechanisms can assist program design in various ways. First, this can help us identify solutions to crime problems. Second, it allows us to capitalize on effects that were once considered unanticipated. Lastly, it helps us produce evaluations and systematic reviews with more valid estimates of the impacts of interventions. This section explains how our framework can assist in this process.

To design effective interventions, we must ground them in two types of theory: intervention and implementation (see Tilley 2004). The former requires a firm understanding of how an intervention should reduce a problem. It should also predict when those effects should appear. Conversely, implementation theory describes how an intervention is to be rolled out. A single intervention can be implemented in multiple ways. Thus, when designing a program, practitioners must understand how the intervention is theoretically supposed to work and how their means of implementation are supposed to produce the intended outcome. While an intervention program could be theoretically sound, the chosen means of implementation could be flawed. If the intervention fails to produce desired effects, it could be abandoned. However, if a different form of implementation were used, desired outcomes could materialize (Tilley 2004).

Theory can also assist practitioners to select the most promising interventions, even if they have not been empirically tested. In fact, having empirical support for a successful intervention in one place is not a strong indicator of success in another context (Cartwright and Hardie 2012). If this same intervention was based on an empirically validated general theory of crime, the evidence about a singular success gives greater confidence that it may work if applied elsewhere (Eck 2017). A strong theory should include a time-course outlining when to expect effects. It will also assist practitioners in determining the inputs required to produce desired outcomes. A better theoretical understanding of an intervention, regardless of its empirical support, will be more helpful in selecting the appropriate strategy.

Returning to our hypothetical police crackdown and alley-gating examples, both interventions have empirical support for their effectiveness in reducing crime (Bowers et al. 2004a; Weisburd and Green 1995). However, neither intervention is necessarily better than the other. Both examples were used above because they are theoretically able to reduce residential burglary via different means. But when assessing the entire time-course of these interventions, practitioners should consider several factors.

For instance, a police crackdown could potentially yield a greater reduction in burglaries. The increased police presence may act as a stronger deterrent to offenders by increasing their likelihood of being caught. However, this strategy also requires a great deal more police resource inputs than alley-gating. A police department may need to increase the number of officers to patrol these areas thoroughly: for the residual deterrence (uncertainty) mechanism to be triggered, the visibility and credibility mechanisms must be sufficiently triggered. This is costly to carry out successfully and the police agency may not be able to carry this out for a long time. Linning and Eck (2018) suggest that implementing a strategy at too low of a dosage could increase crime by creating more opportunities. Moreover, some research suggests that benefits post-intervention decay quite quickly (Sorg et al. 2013). Thus, the overall program would require sufficient resource investments that would also need to be re-applied shortly thereafter with rotating crackdowns (Sherman 1990). This approach also requires continual department and officer buy-in. It must compete with any new strategy that becomes popular and potentially less costly regardless of whether it works.

On the other hand, the alley-gating policy also requires much buy-in but from different interest groups. The most successful cases of alley-gating involve implementing the gates on private property, not public right of way (Merton Metropolitan Police 2015; Sidebottom et al. 2017). Thus, practitioners must obtain consent from all residents and business owners who may be affected by their installation. Furthermore, unlike the police crackdowns funded through tax dollars, alley-gates can have high upfront costs that may be incurred to residents (Merton

Metropolitan Police 2015). However, once in place, they require little economic investment other than minor upkeep. Thus, even if the gates are more economically favorable in the long run, the up-front costs and community consultation required may make their installation more difficult to implement than a crackdown.

Although many interventions can have empirical support for their effectiveness, such evidence is not necessarily an indicator of high external validity. This discussion highlights some of a strategy's components that should be considered to determine whether a strategy might work in various contexts. To do this most effectively practitioners need a strong theoretical understanding of the intervention being considered (Eck 2017). Cartwright and Hardie (2012) refer to these components as support factors. They liken the determination of an effective strategy to baking a cake. Without every ingredient, the final product will not be the desired cake. Unfortunately, we are often unaware of all the ingredients and sometimes add unnecessary ones. Our framework can assist practitioners to identify essential ingredients. Failing to consider an important support factor (or ingredient) can be detrimental to the success of a program. For instance, in neighborhoods with high resident turnover or where residents shared number combinations of alley-gate keypads, the intervention was less effective (Sidebottom et al. 2017). Similarly, alley-gates in areas with high student populations also reported observing many gates left propped open (see Millie and Hough 2004). Thus, adherence to the intervention by residents is a key ingredient that must be considered to successfully reduce crime.

In addition, Cartwright and Hardie (2012) argue that support factors usually must be added in a sequence to properly function. By identifying the step-by-step process (or time-course) of the intervention, we can identify many of the required support factors needed for success. For alley-gating, this implementation process is outlined by Sidebottom et al. (2017). It starts by consulting with local residents and owners as well as police. If interest is present, this can be followed by determining the legal status or ownership of the alleys. Next, consent to carry out the process must be obtained from the relevant property owners. Once this occurs, site preparation, construction, and installation can take place (see Sidebottom et al. 2017). Many steps occur in the time-course prior to the intervention actually starting. Yet, as noted, crime-reducing effects can be observed during these stages and thus should be considered. This can only be attained if practitioners have a firm grasp of the theoretical sequence and support factors needed to properly trigger the mechanisms of change. And these support factors can vary with every case.

EVALUATING INTERVENTIONS

Considering the entire time-course also has implications for evaluation that quantify the size and significance of any crime reduction that may come from an intervention. A lack of understanding of the potential role of unmeasured temporal effects in confounding outcome analysis is perhaps the implication of most concern. Figures 4a and 4b below demonstrate two scenarios in which an unrecognized temporal effect can lead to an inaccurate estimation of outcome. In the first case, the lack of detection of an initial backfire effect could lead to a misrepresentation of effectiveness in a simple quasi-experimental design because the experimental ‘after’ period includes the time before it had reached its desired performance level (figure 4a). In the second, under-estimation occurs because the ‘after’ period excludes anticipatory benefits and includes them in the pre-intervention period (figure 4b).

<INSERT FIGURES 4a & 4b ABOUT HERE>

These measurement issues raise definitional considerations for prevention interventions, specifically, when does it begin and what actions are part of that intervention? In this example, does the intervention begin once the alleys have gates blocking them? Is the intervention the presence of the gates? Or does the intervention begin at the first moment potential offenders suspect changes to the alleys are imminent? And is the intervention a combination of the gates and knowledge about the gates within the surrounding community? As gates cannot be instantaneously installed without warning, the pre-gate publicity must be considered as part of the intervention.

The central point here is that people involved in planning prevention programs need to consider which parts are separable and which are not. A useful mental exercise might be to consider whether it is possible to compare multiple forms of the program in an evaluation. Comparing alley-gating with and without publicity seems impossible, so the publicity is integral to the intervention and anticipatory benefits and backfire must be considered. In contrast, it is possible to compare crackdowns with and without community notification. In this circumstance, the planners need to be explicit about their plans for publicity. Just as important, when evaluators describe their results, they need to be explicit about the publicity ingredients. The same style of thinking should go into all the pieces of an intervention.

If details at all of the stages in the time-course are present, we will be able to evaluate program effectiveness more completely. To understand the importance of our proposal, compare the following hypothetical evaluations (see table 1). In case A, the researcher is able to measure outcomes for anticipatory benefits (0.10), main immediate benefits (0.70), and residual benefits (0.05). She then sums these to provide an overall estimate of the total benefits from the intervention (0.85). In case B, the researcher only examines the main immediate benefits (0.65), which also serve as the total benefit estimate. It should be readily apparent that the two estimates of total benefits would only be the same if there were no anticipatory benefits and no residual benefits in the second study. In any other circumstance, estimates from case A would be more valid than estimates from case B. We do not know, in this example, whether the intervention in case A worked better than in case B, or whether they are about equally effective, or even if case B really produced superior results.

<INSERT TABLE 1 ABOUT HERE>

This simple comparison demonstrates that considering the time-course is likely to improve the quality of evaluations, whether a randomized trial or a strong quasi-experiment. How does taking the time-course of interventions improve the quality of systematic reviews and meta-analysis? In the hypothetical example depicted in Table 1, a reviewer has discovered four studies on the intervention (we will assume they are of equal size and quality in all other regards). Evaluations A and B are from the first two scenarios, just described. C and D are two additional evaluations. Note that evaluation D detected a possible backfire effect (negative value in the residual column). It is clear that the total impacts described in these four studies (last column) will be drawing upon extremely different bodies of knowledge and only in evaluation A will something like a complete estimate be available.

Now look at the columns. Here the reviewer might estimate the effect sizes (at the base of the columns). Only in the "immediate" column would the reviewer be able to draw upon all four studies. If the reviewer did not take into consideration the absence of information, their effect size for the total impact of the four studies would be biased (unless there were no anticipatory or residual outcomes). Because the absence of information from studies is a form of measurement error in the calculation of the total effect size, the total impact estimate would also be more uncertain. The reviewer would only know this if she looked for outcome measurements along the time-courses of each experiment.

Table 1 illustrates that if the studies reviewed contain outcome estimates at each stage of the time course, then a reviewer can estimate when the benefits from studies are likely to occur. Instead of having a single overall estimate, the reviewer can show effect sizes throughout the time-course. Consequently, reviewers could report that some interventions have the bulk of their benefit long after the intervention, others have their greatest impact right after the intervention, and some create their most positive results before the formal unveiling of the intervention. Confidence intervals at each stage would show the uncertainty associated with outcomes. Selectively combining different interventions with different payout schedules might produce more sustained benefits.

Farrell et al. (2005) provide the only known crime prevention study that compares estimates at various stages of our proposed time-course. Specifically, when evaluating the impact of a burglary reduction scheme, the authors estimated that 308 burglaries were prevented as part of the immediate intervention over a two-year period. However, they also found that an additional 419 burglaries were prevented as a consequence of anticipatory benefits. Similarly, 37 burglaries were believed to be prevented via diffusion of benefits. Though these additional 37 crimes are attributed to spatial effects, they provide yet another layer of effects beyond the immediate intervention outcome. When considering these effects together, the program reduced 764 burglaries overall. This shows a much greater impact beyond the time-isolated estimate of 308 offenses.

These examples reinforce our need to be attentive to research designs. Different research designs are variable in the degree to which they exhibit internal and external validity. Simple before- and after-designs are likely to be most inadequate, whether or not they also have controls (i.e., are non-equivalent control group designs). Such designs can lead to the miscalculation of impacts when there is ambiguity over when a program begins but the evaluators assume that a fixed after period sufficiently captures the entirety of the impact period. The introduction of controls is unlikely to guard against this type of failure because the post-intervention period for controls and treatments are mis-specified. Further, randomized controlled trials too will fall victim to these issues. Though the use of random assignment improves our ability to assess treatment effects relative to the control, failing to consider the time-course when completing observations can lead to over- or under-estimations of the intervention's true impact.

Thus, longitudinal research designs may be more useful when temporal effects are possible. Whilst this is not a particularly surprising revelation, several points are useful to raise here. First, studies that track impact over extended time-courses (for a time series both before and after 'start' dates) are much more likely to capture the true 'policy life course' effect of a particular intervention. When longitudinal evaluations measure outcomes at many time periods of short durations (e.g., months rather than years), then the evaluator can present findings under alternative assumptions (e.g., assume the program had no warning or publicity versus assume the program became known to the public four months prior to its official start date).

Second, longitudinal designs can track the length of the outcomes. Various forms of pre-post designs cannot determine when the intervention ceases to have much effect: they show when the evaluator stopped looking. In a randomized controlled study of the impact of intervening with landlords on crime at places, for example, Eck and Wartell (1998) showed that 46 percent of the impact occurred in the first six months following the interventions with the remainder of the impact trickled in over the next 24 months. Here, a randomized control trial using a pre-post design would have dramatically cut the estimated: instead of a 60 percent reduction in crime, it would have found less than a 30 percent drop in the treated places relative to their controls. This was a human resource type of intervention, like a crackdown. The timescales over which a physically fitted measure (e.g., an alley-gate) is likely to offer reduction should be much longer. This is almost never measured in outcome evaluations quite possibly due to the resource demands of conducting such assessments. A longitudinal design can help estimate when the impact of an intervention is likely to be zero, even if that point in time is outside the measurement range. To do this, however, requires far more measurements than a standard pre-post design provides.

Third, longitudinal designs have far more scope in accommodating the temporal subtleties of the delivery of the intervention itself. In reality, few interventions exist that are delivered in their entirety on a specific date. More likely is that the intensity of intervention ebbs and flows over a time-course. It has been argued elsewhere that tracking the relationship between trends in realized outputs over time and crime reduction can provide another form of evidence that it was indeed the intervention that lead to the outcome observed (Bowers et al. 2004b). Adding data on the actual distribution of inputs and outputs is only an option with statistical approaches that can accommodate such time-varying independent variables. Including them can also give the evaluator clues as to crime trends that are related to the actual outputs of the intervention itself and those that are possibility related to some of the other temporal effects listed above.

A further point is that tracking what is actually done matters for understanding the process of an intervention and for its potential generalization to other contexts. This links back to the principles of the EMMIE framework. Without really considering how an intervention works and the context in which it operates it is difficult to disentangle what might be called the ‘stylized’ effects of the intervention itself from temporal (and spatial) effects caused by the particular incarnation of that intervention. Over- or under-estimation of the true effect of the specific intervention outputs has an impact not only on crime reduction estimation, but also on cost-effectiveness calculations (Farrell et al. 2005). Further, any exercises that synthesize crime reduction results across several evaluations will not be immune to these potential problems of over- or under- estimation of effect size. Meta-analyses often draw on primary evaluation exercises with a wide variety of different experimental designs. As demonstrated above, if some of these consider temporal effects in their evaluations and some do not, the true effect size of the intervention itself is unlikely to be reliably estimated in the mean effect size.

Lastly, even if evaluators cannot use a time series design to measure intervention effectiveness, they should still consider their evaluation in longitudinal terms. For instance, if resource constraints only allow you to collect one pre- and one post-test measure, understanding the intervention's time-course is crucial. Suppose the strategy is effective, but only after an initial two-month phase of initial backfire (recall figure 4a). If evaluators take their post-test measurement at the end of the first month following the intervention, it will appear largely ineffective. However, if evaluators consider the time-course of the intervention and expect this temporary backfire, taking that single measurement three months following the intervention start date gives a far more accurate indication of its effectiveness.

CONCLUSIONS

In this paper we unite many previously identified crime prevention effects, namely anticipatory benefits, residual deterrence, and initial backfire (Linning and Eck 2018; Sherman 1990; Smith et al. 2002). Considering them within a single framework would benefit the design of experiments and evaluations. We suspect that most prevention strategies will produce only some, or maybe none, of these effects. Uncertainty over which effects prevention strategies generate creates potential threats to valid policy recommendations. Consequently, these temporal effects should always be considered, and evaluations should be designed to account for them. In addition to the consequences for assessing evidence of effectiveness, practitioners need to be cognizant of these effects so they can enhance positive effects and limit negative ones.

To predict whether various temporal effects will occur depends on practitioners' understanding of the causal mechanisms the intervention is designed to trigger. Much of the current literature indicates that we have a weak understanding of these mechanisms. New advancements such as the EMMIE framework (Johnson et al. 2015) have highlighted the importance of considering more than just intervention effect sizes. Identifying the mechanism that is supposed to reduce crime is not enough. We also need to understand what contexts are needed to trigger it in a beneficial way. Practitioners also need to map out the economic costs of an intervention including what implementation components it demands.

Multiple mechanisms can be triggered within a single intervention. Moreover, their effects can be triggered at different points in time. To be most effective, practitioners must consider changes in inputs, resources, and implementation requirements at various stages during the intervention's time-course. When multiple mechanisms are triggered, we must also recognize whether they are temporally dependent on one another. We proposed the concept of a *mechanism cascade* to point this out. In our police crackdown example, the preventative effects of triggering the offender uncertainty mechanism will not occur unless the officer visibility and credibility mechanisms are triggered beforehand. That said, some mechanisms operate independently and such timing is not required. However, this must be considered prior to implementation to reap the greatest benefits. If this does not happen, an otherwise effective intervention could fail (and may even be mistakenly abandoned) because it was not implemented correctly. In many cases, practitioners do have the resources to carry out an intervention properly. These failures may simply occur because of insufficient knowledge, thought, and planning going into the intervention's design. In some instances, practitioners might luck out if their mechanisms happen to work independently. But we believe that using our time-course framework will leave much guessing out of the planning stages and increase one's likelihood of successful implementation.

A better understanding of when to expect effects also improves program evaluation and policy making. Despite making some seemingly unoriginal claims—for instance, the use of longitudinal evaluations of interventions has been suggested before—our time-course framework can better inform resource-constrained evaluations. If evaluators plot out when particular effects are expected, they can more strategically select when is best to collect their pre- and post-test measurements. Similarly, mapping out the time-course of an intervention can help practitioners better understand their interventions and avoid adverse backfires. Knowing that prevention policies can fluctuate in how they deliver outcomes, policymakers should design their prevention strategies to accentuate desirable results and dampen undesirable results. So, for example, if it is feasible to create anticipatory benefits through careful communications, they should do so, if the costs of creating anticipatory benefits is low. Or, for example, if initial backfire is possible, steps should be taken to minimize or eliminate it. Extra police resources can be deployed to suppress a temporary crime increase while awaiting the crime reduction effect. This will make the intervention even more effective overall.

If interventions have time-courses that produce multiple effects and mechanism cascades, then evaluators need to be very explicit about their intervention and its mechanisms. If readers do not know the details of the intervention, they will have difficulty synthesizing and interpreting evidence and there will be far greater variation in evaluation outcomes of the same program.

Indeed, it is quite possible that much of the variation we see in systematic reviews and meta-analysis is due to grouping together programs that are quite different except for their names.

The time-course framework we propose provides an advance in crime prevention experiments, evaluations, and systematic reviews. Although we used place-based examples as illustrations, it can be applied to other interventions. For instance, evaluations of correctional programs would benefit from considering the time-course of treatment programs. Offenders may improve their behavior with hopes to gain entry into programs. Conversely, backfires (e.g., prison misconduct) may occur after treatment as a consequence of increasing contact among inmates. As such, we believe this framework will help practitioners, evaluators, and policy-makers expand their thinking and improve program design. Multiple important discoveries have been made in this area and the consolidation of these effects into a single framework is both theoretically possible and likely of benefit to our crime prevention efforts.

References

Bowers, K. J., & Johnson, S. D. (2004). Who commits near repeats? A test of the boost explanation. *Western Criminology Review*, *5*(3), 12-24.

Bowers, K. J., Johnson, S. D., & Hirschfield, A. F. (2004a). Closing off opportunities for crime: An evaluation of alley-gating. *European Journal on Criminal Policy and Research*, *10*(4), 285-308.

Bowers, K.J., Johnson, S.D. & Hirschfield, A. (2004b). The measurement of crime prevention intensity and its impact on levels of crime. *British Journal of Criminology, 44*(3), 1-22.

Burrows, J., & Heal, K. (1980). Police car security campaigns. In R.V. Clarke & P. Mayhew (eds.), *Designing out Crime*. London, UK: Her Majesty's Stationery Office.

Clarke, R. V. (1995). Situational crime prevention. In M. Tonry & D. P. Farrington (eds.), *Building a Safer Society: Strategic Approaches to Crime Prevention*, vol. 19, 91-150. University of Chicago Press.

Clarke, R. V., & Weisburd, D. (1994). Diffusion of crime control benefits: Observations on the reverse of displacement. *Crime Prevention Studies, 2*, 165-184.

Eck, J. E. (2017). Some solutions to the evidence-based crime prevention problem. In J. Knuttson and L. Tompson (eds.), *Advances in Evidence-Based Policing* (pp. 45-63). New York, NY: Routledge.

Eck, J. E. (2010) Policy is in the details: Using external validity to help policy makers, *Criminology & Public Policy* 9(4): 859-866.

Eck, J. E. & Wartell, J. (1998). Improving the management of rental properties with drug problems: A randomized experiment. In L. G. Mazerolle & J. Roehl (Eds.), *Civil Remedies and Crime Prevention*, Crime Prevention Studies (Vol. 9, pp. 161–185). Monsey, NY: Criminal Justice Press.

Farrell, G., Bowers, K.J., & Johnson, S.D. (2005). Cost-benefit analysis for crime science: Making cost-benefit analysis useful through a portfolio of outcomes. In M. Smith & N. Tilley (eds.), *Crime Science: New Approaches to Preventing and Detecting Crime* (pp. 56-84). Portland, OR: Willan Publishing.

Gambetta, D. (1998). Concatenations of mechanisms. In P. Hedstrom & R. Swedberg (eds.), *Social mechanisms: An analytical approach to social theory*. Cambridge University Press.

Guerette, R. T., & Bowers, K. J. (2009). Assessing the extent of crime displacement and diffusion of benefits: A review of situational crime prevention evaluations. *Criminology*, *47*(4), 1331-1368.

Johnson, S. D. & Bowers, K. (2005) Using Publicity for Preventive Purposes, in Nick Tilley (Ed) Handbook of Crime Prevention: Theory, Policy and Practice (first edition). London: Willan

Johnson, S. D., & Bowers, K. J. (2004). The burglary as clue to the future: The beginnings of prospective hot-spotting. *European Journal of Criminology, 1*(2), 237-255.

Johnson, S. D., Tilley, N., & Bowers, K. J. (2015). Introducing EMMIE: an evidence rating scale to encourage mixed-method crime prevention synthesis reviews. *Journal of Experimental Criminology*, *11*(3), 459-473.

Laycock, G. (1991). Operation identification or the power of publicity? *Security Journal, 2*(2), 67-72.

Linning, S. J., & Eck, J. E. (2018). Weak intervention backfire and criminal hormesis: Why some otherwise effective crime prevention interventions can fail at low doses. *The British Journal of Criminology, 58*(2), 309-331.

Merton Metropolitan Police (2015). A guide to alleygates. Retrieved from: https://www.bexley.gov.uk/sites/bexley-cms/files/A-guide-to-Alleygates.pdf

Miller, J. G., & Hoelter, H. H. (1979). Prepared testimony: Oversight on Scared Straight. *Washington, DC: Government Printing Office*.

Millie, A. & Hough, M. (2004). *Assessing the impact of the reducing burglary initiative in southern England and Wales* (Home Office Online Report 42/04). London, UK: Home Office.

Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London, UK: Sage.

Riley, D. (1980). An evaluation of a campaign to reduce car thefts. In R.V. Clarke & P. Mayhew (eds.), *Designing out Crime*. London, UK: Her Majesty's Stationery Office.

Ross, H. L. (1981). *Deterring the drinking driver: Legal policy and social control*. Lexington, Mass.; Health.

Sherman, L. W. (1990). Police crackdowns: Initial and residual deterrence. *Crime and Justice*, *12*, 1-48.

Sidebottom, A., Tompson, L., Thornton, A., Bullock, K., Tilley, N., Bowers, K., & Johnson, S. D. (2017). Gating alleys to reduce crime: A meta-analysis and realist synthesis. *Justice Quarterly*, 1-32.

Smith, M. J., Clarke, R. V., & Pease, K. (2002). Anticipatory benefits in crime prevention. *Crime Prevention Studies*, *13*, 71-88.

Sorg, E. T., Haberman, C. P., Ratcliffe, J. H., & Groff, E. R. (2013). Foot patrol in violent crime hot spots: The longitudinal impact of deterrence and posttreatment effects of displacement. *Criminology*, *51*(1), 65-101.

Tilley, N. (2004). Applying theory-driven evaluation to the British Crime Reduction Programme: The theories of the programme and its evaluations. *Criminal Justice, 4*, 255-276.

Tompson, L., Belur, J., Thornton, A., Bowers, K., Johnson, S., Sidebottom, A., Tilley, N. and Laycock, G. (2019). Taking stock of systematic reviews in crime reduction: An evidence appraisal using the EMMIE framework. Under Review.

Weisburd, D., & Green, L. (1995). Policing drug hot-spots: The Jersey City drug market analysis experiment. *Justice Quarterly, 12*, 711-735.

Figure 1. Temporal Effects across Offending Timeline

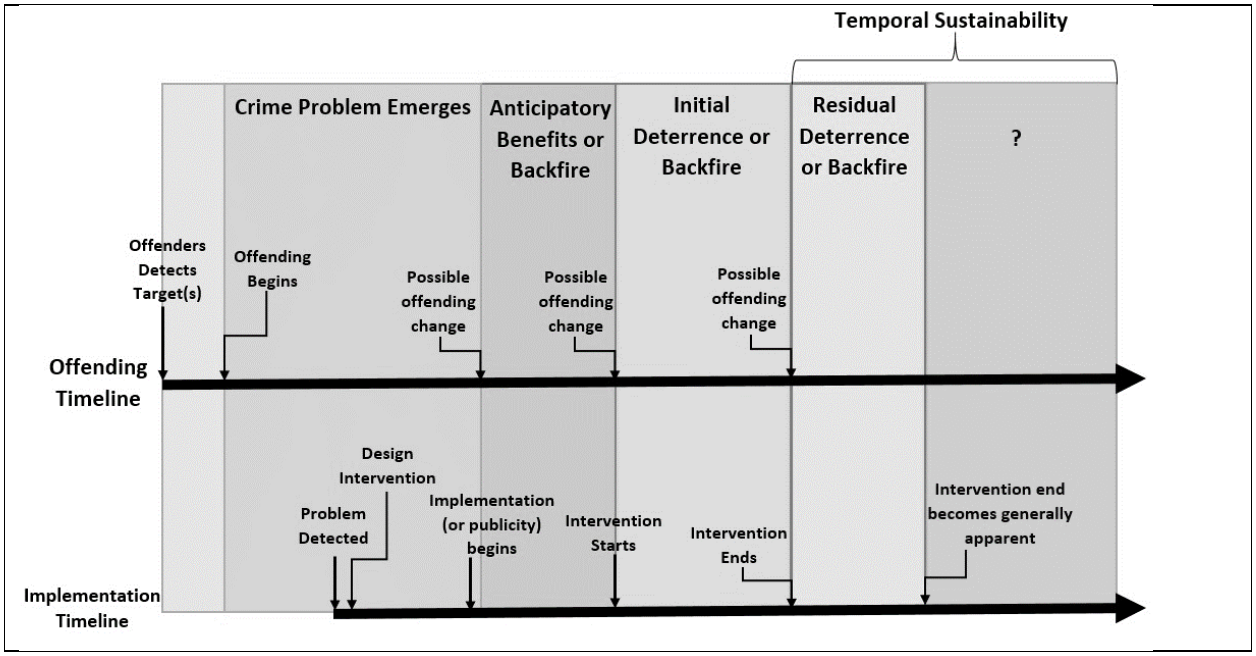

Figure 2. Crime prevention time-course including all EMMIE dimensions, police crackdown strategy

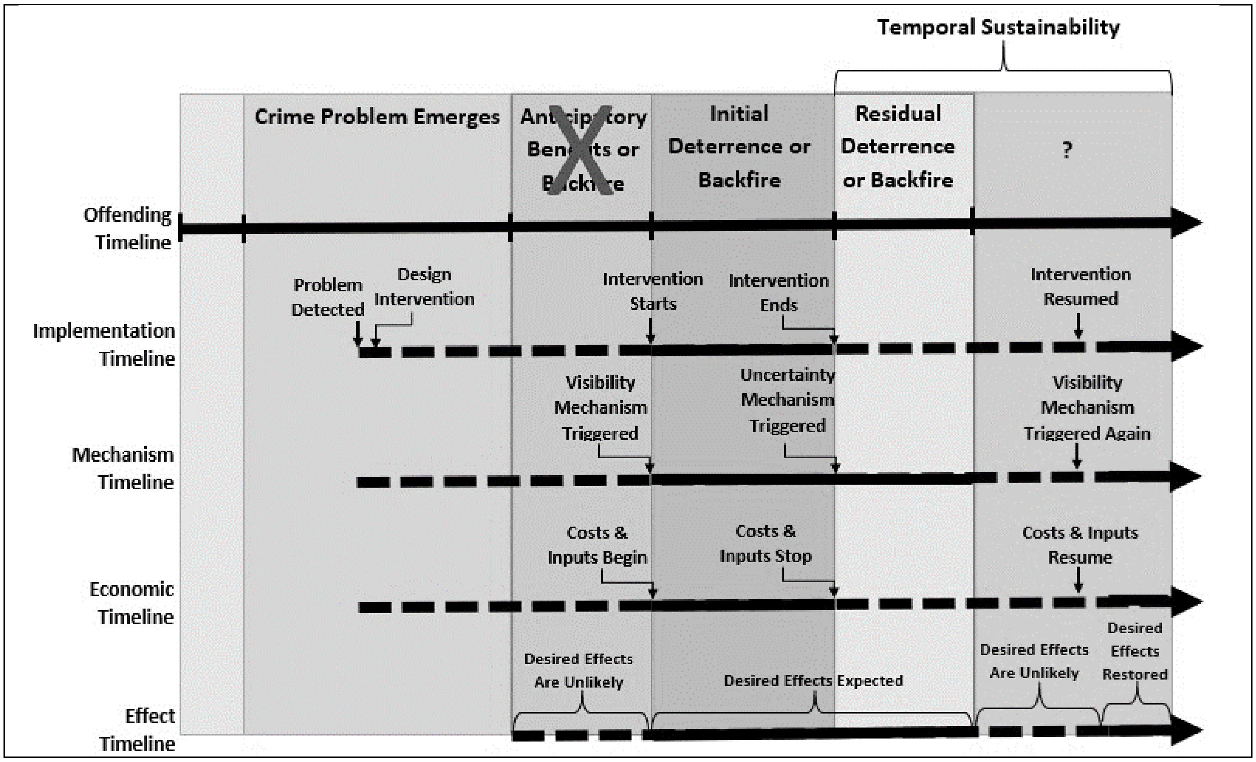

Figure 3. Crime prevention time-course including all EMMIE dimensions, alley-gating strategy

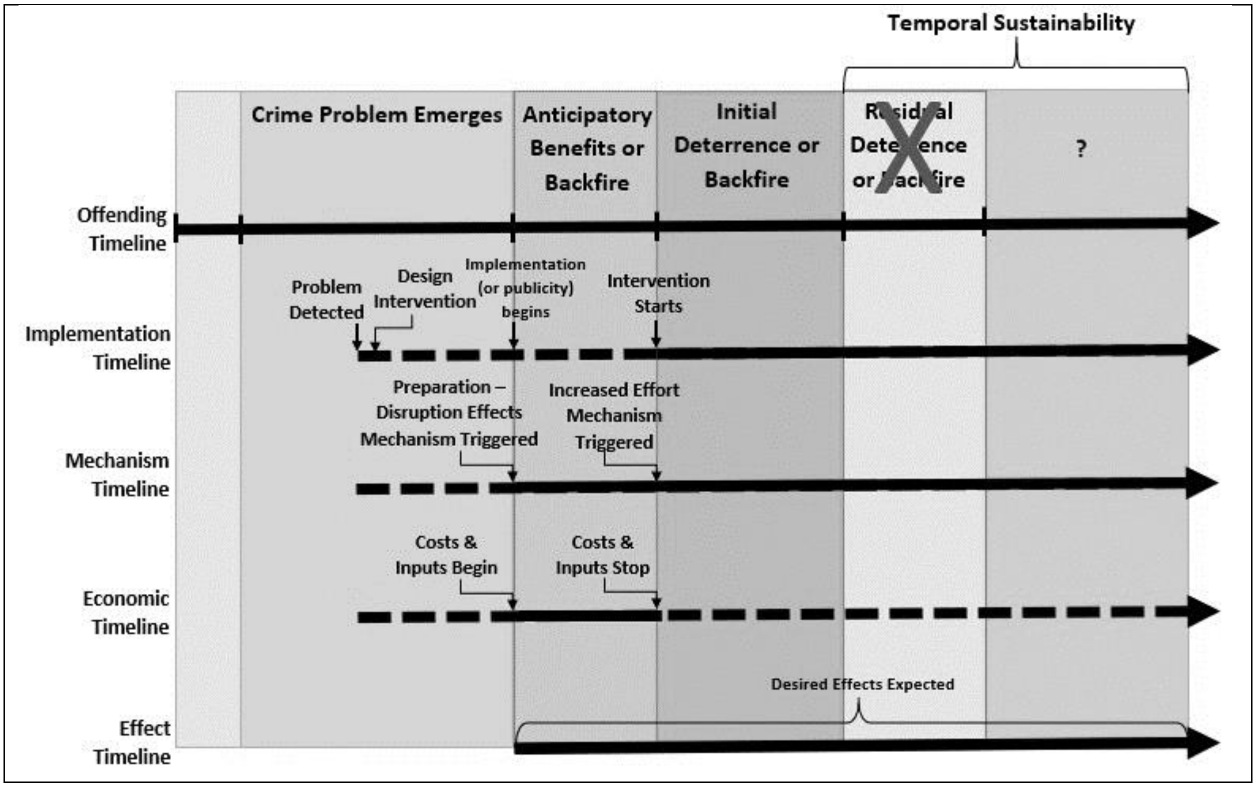

Figure 4a. The under-estimation of effectiveness, initial backfire

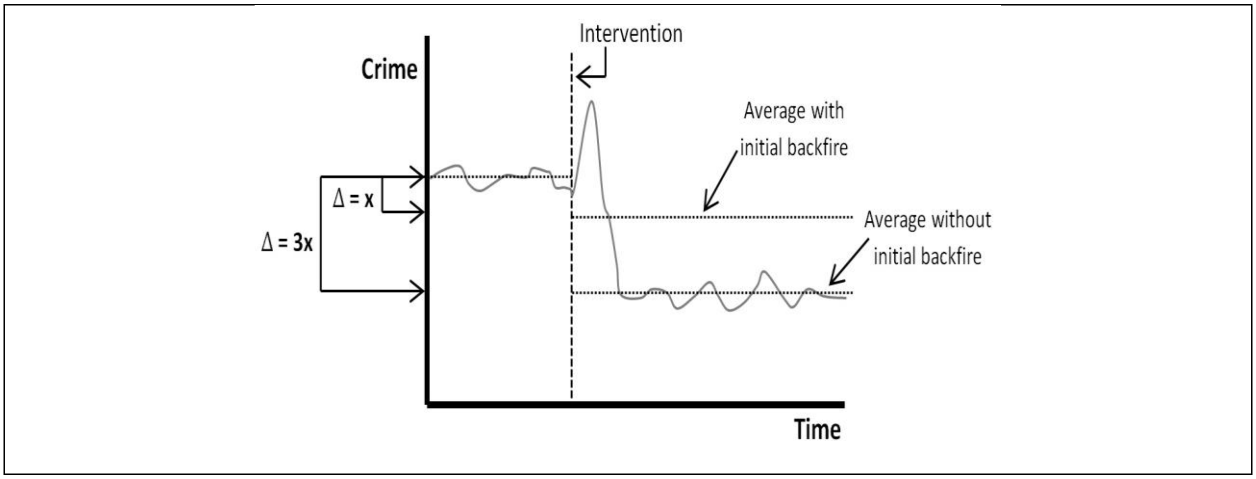

Figure 4b. The over-estimation of effectiveness, anticipatory benefits

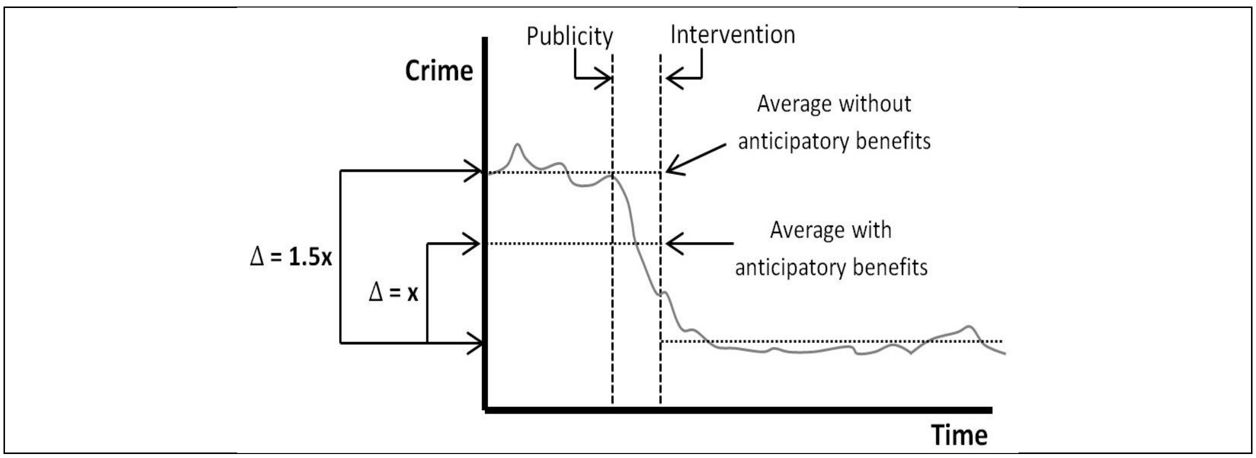

Table 1. A hypothetical review of four studies measuring different parts of the time-course of the same intervention.

Evaluation	Measured Outcomes*			
	Anticipatory	Immediate	Residual	Total Impact
A	0.10	0.70	0.05	0.85
B	Not measured	0.65	Not measured	0.65
C	0.07	0.50	Not measured	0.57
D	Not measured	0.25	-0.10	0.15
Review estimate**	0.09	0.53	-0.03	0.59

* The numbers are hypothetical and do not represent any specific metric. Rather they are used to illustrate how any standardized measure could be used.

** Mean of measured outcomes, assuming studies of equal size. In reality, these studies would be weighted differently.